

Divide and Conquer: An Incremental Sparsity Promoting Compressive Sampling Approach for Polynomial Chaos Expansions

Negin Alemazkoor, Hadi Meidani*

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.

Abstract

This paper introduces an efficient sparse recovery approach for Polynomial Chaos (PC) expansions, which promotes the sparsity by breaking the dimensionality of the problem. The proposed algorithm incrementally explores sub-dimensional expansions for a sparser recovery, and shows success when removal of uninfluential parameters that results in a lower coherence for measurement matrix, allows for a higher order and/or sparser expansion to be recovered. The incremental algorithm effectively searches for the sparsest PC approximation, and not only can it decrease the prediction error, it can also reduce the dimensionality of PCE model. Four numerical examples are provided to demonstrate the validity of the proposed approach. The results from these examples show that the incremental algorithm substantially outperforms conventional compressive sampling approaches for PCE, in terms of both solution sparsity and prediction error.

Keywords: Compressive Sampling, Polynomial Chaos Expansion, Uncertainty Quantification, Sparse Approximation, Legendre Polynomials.

1. Introduction

To reliably predict the behavior of a physical system of interest, it's vital to accurately model the interaction between the system's inputs and outcomes, and account for the input uncertainty and its impact on the outcomes or the quantities of interests. The uncertain inputs to the model can include uncertain initial or boundary conditions, uncertain parameters, etc., which are often modeled as random variables. Uncertainty quantification (UQ) approaches are used to model the propagation of random input variables through the model and quantify the uncertainty in the outputs. The most commonly known and used UQ approach is the Monte Carlo approach, which despite its simple and robust applicability, suffers from a poor convergence rate of $\mathcal{O}(M^{-\frac{1}{2}})$, where M is the number of samples. Although extensive research has been done and several modifications to the original Monte Carlo approach have been proposed to improve this convergence rate, e.g. importance sampling [1], quasi-Monte Carlo [2], multilevel Monte Carlo [3]) these sample-based approaches still suffer from slow rate of convergence. Therefore, there has been a growing interest in exploring alternative numerical approaches for UQ.

*Corresponding author

Email address: meidani@illinois.edu (Hadi Meidani)

Stochastic spectral approaches, as one alternative, have been shown to offer accelerated convergence and speed-up over Monte Carlo approaches, thereby being recognized as a promising solution to a variety of engineering problems especially when the dimensionality of parameter space is not large. In these methods, the random output is represented through a spectral expansion on specific orthogonal basis functions such as generalized polynomial chaos (PC) [4, 5]. There are two distinguished classes of methods to evaluate the coefficients of the polynomial expansion, stochastic Galerkin and stochastic collocation methods. Stochastic Galerkin method is an extension of Galerkin approach, in which continuous operator problems such as differential equations are converted to a system of equations. Similar to Galerkin scheme, stochastic Galerkin leads to a K -coupled equation system, where K is the number of coefficients to be estimated in polynomial chaos expansion (PCE), making it an intrusive approach. When the original problem has a complex nonlinear form, deriving the coupled equations can be cumbersome or even impossible.

Alternatively, in non-intrusive stochastic collocation methods, the coefficients are estimated either by solving a least square problem to fit a set of sample responses or by recasting the coefficients as integrals and approximating the integrals using sampling, quadrature, or sparse grid approaches. Despite the Galerkin method, the complexity of the original problem does not impact the applicability of collocation methods. However, due to the curse of dimensionality, when the number of uncertain input parameters is large, even when efficient techniques such as sparse grid [6–8] are used, prohibitively large number of sample points are required for accurate estimation of chaos coefficients. Adaptive methods such as adaptive sparse grid [9, 10] and functional ANOVA decomposition [11, 12] have been developed in order to reduce the required number of samples by identifying the unimportant dimensions and placing fewer sample points in those dimensions.

Recently, additional efforts have been made to explore the regularity of the solution using a small number of samples when the solution is sparse and its approximated PCE has only few terms. Specifically, compressive sampling was used in [13–15] as a non-intrusive non-adapted approach to approximate the sparse solution of stochastic problems. Compressive sampling was first initiated in the field of signal and image reconstruction [16–18]. Conventionally, based on the distinguished Shannon’s theorem, sufficient sampling rate was considered to be larger than twice the maximum frequency in the signal [19]. Moreover, the fundamentals of linear algebra also impose the number of samples to be equal or larger than the dimensionality of the signal to ensure the reconstruction. However, compressive sampling techniques allowed reconstruction of sparse signals and images from incomplete measurements, i.e. a small set of samples, and effectively solving an underdetermined linear system [16–18].

In non-intrusive estimation of PCE, samples are acquired from an actual experimentation on the behavior of the system or from numerical simulations of a high-fidelity model for the system. Both experimentation and simulation of complex systems can be very expensive, calling for efforts to reduce the number of required samples in non-intrusive estimation procedures. In [13, 14], compressive sampling was successfully applied in estimating the chaos coefficients where the number of samples was significantly smaller than the number of unknown chaos coefficients. However, this success heavily depends on two conditions: (a) the solution of stochastic problem should be in fact sufficiently sparse, and (b) the measurement matrix which includes random evaluations of polynomial bases should be sufficiently incoherent [13]. In [20], the authors focused on improving the first condition, i.e. the actual sparsity of the solution, by rotating the random inputs using the active subspace method [21]. As a result of this rotation, only a few influential bases will effectively participate in the spectral representation of the quantity of interest (QoI), making the target representation sparser. However, as the authors admit, the coordinate rotation has negligible

impact on the coherence of the measurement matrix. This means that the second condition for a successful sparse recovery is not improved. That is, no matter how effectively one can transform the coordinate system of the random inputs to enable a sparser *target* solution, the attempts to recover that target solution using the available samples will still be in vain if the measurement matrix is highly coherent. Our proposed approach tries to be aware of this second condition on the coherence. Specifically, instead of rotating the uncertainty sources, it concerns selecting for the polynomial expansion the right order and dimension, which are two factors that directly impact the coherence of the measurement matrix. This selection is done given a fixed sample set and basis type and an inaccuracy tolerance level. In what follows, we elaborate on how each one of these two factors can have implications on the success of sparse recovery.

First, an increase in the total order for a PCE increases the number of columns in measurement matrix which makes it undesirably coherent. Therefore, the expansion order is usually chosen to be small, especially when the dimension of parameter space is large. This will deteriorate the accuracy if the exact solution is in fact of a higher order. In addition to the expansion order, an increase in the dimensionality also directly inflicts large coherence on measurement matrix, thereby degrading the accuracy of compressive sampling solution. In [22], an adaptive algorithm is introduced to identify and include only the important polynomial coordinates with the aim of avoiding large coherence for the measurement matrix. Their algorithm begins with an initial basis set that includes all the dimensions but have a low total order, typically equal to one. At each step of the algorithm, several anisotropic trial bases of higher orders in different dimensions are considered and the best one is selected to be included the basis set according to a cross-validation error criterion.

In this work, we introduce a novel incremental compressive sampling algorithm, which considers the detrimental impact of coherence on solution's sparsity and accuracy, and searches for the sparsest solution by exploring through combinations of dimensionality and order of a PCE. Compared to the adaptive algorithm proposed in [22], our criterion for selecting new bases or stopping the algorithm is the sparsity of solutions calculated at each iterations. Also, the initial basis set in our algorithm can include as few as one dimension. This will remove the possibility of facing a large coherence hurdle at the initial step when high dimensional problems are concerned. This paper is organized as follows. Section 2 presents general concepts in compressive sampling and its theoretical background. In Section 3, we introduce our incremental compressive sampling algorithm along with relevant theoretical supports. Finally, Section 4 includes numerical examples and discussions.

2. Compressive sampling in spectral uncertainty analysis

2.1. Spectral uncertainty analysis

Spectral uncertainty analysis methods aim to efficiently propagate the input uncertainty in numerical models. These methods are constituted based on expansions on orthogonal polynomial bases and have been proved to have spectral convergence [4, 5, 23–25]. Formally, let $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the event space, \mathcal{F} is a σ -field over Ω , and \mathbb{P} is the probability measure on (Ω, \mathcal{F}) . Also, let $\Xi = (\Xi_1, \dots, \Xi_d)$ be a d -dimensional random vector with mutually independent components, joint density function, $f_{\Xi}(\xi)$, or in short $\rho(\xi)$, and finite moments on Ω . Then any square-integrable random quantity of interest $u(\Xi) : \Omega \rightarrow \mathbb{R}$ can be represented as:

$$u(\Xi) = \sum_{\alpha \in \mathbb{N}_0^d} c_{\alpha} \psi_{\alpha}(\Xi), \quad (1)$$

where $\{\psi_{\alpha}\}_{\alpha \in \mathbb{N}_0^d}$ is the set of orthogonal basis functions satisfying

$$\int_{\Omega} \psi_{\mathbf{n}}(\boldsymbol{\xi}) \psi_{\mathbf{m}}(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} = \gamma_{\mathbf{n}} \delta_{\mathbf{mn}}, \quad \mathbf{m}, \mathbf{n} \in \mathbb{N}_0^d, \quad (2)$$

where $\gamma_{\mathbf{n}}$ is the normalization factor and $\delta_{\mathbf{mn}}$ is the delta function.

For computational sake, finite order truncation of spectral expansion is typically used to approximate $u(\boldsymbol{\Xi})$:

$$u_k(\boldsymbol{\Xi}) := \sum_{\alpha \in \Lambda_{d,k}} c_{\alpha} \psi_{\alpha}(\boldsymbol{\Xi}), \quad (3)$$

where $\Lambda_{d,k}$ is the set of multi-indices defined as

$$\Lambda_{d,k} := \{\alpha \in \mathbb{N}_0^d : \|\alpha\|_1 \leq k\}. \quad (4)$$

The cardinality of $\Lambda_{d,p}$, i.e., the number of expansion terms, here denoted by K , is a function of d and k according to

$$K := |\Lambda_{d,k}| = \frac{(k+d)!}{k!d!}. \quad (5)$$

Given this setting, u_k approximates u in a proper sense and is referred to as the k -th degree PC approximation of u .

Each of the K coefficients involved in u_k definition can be exactly calculated by projecting u onto the associated basis function:

$$c_{\alpha} = \frac{1}{\gamma_{\alpha}} \int_{\Omega} u(\boldsymbol{\xi}) \psi_{\alpha}(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) d\boldsymbol{\xi}. \quad (6)$$

The exact calculation of above integral is a non-trivial and usually impossible task. Alternatively, the integral can be approximated using sampling, cubature rule or sparse grid techniques. However, when d is large, even in efficient approaches such as sparse grid [6–8, 26], the required number of samples is large. For instance, in Clenshaw-Curtis sparse grid, the number of samples need to be about 2^k times the number of expansion coefficients [7]. Acquiring such a large sample set can be infeasible, due to limitations in experiments for empirical samples or limitations in computational resources for the case where samples are obtained from high-fidelity simulations.

As briefly mentioned in Section 1, linear regression can also be used to approximate the expansion coefficients. Regression requires the number of collocation points to be at least equal to K , the number of expansion terms. Although regression techniques significantly alleviate the prohibitive requirement on the number of samples, there has been concerns regarding its accuracy [5]. Therefore, it is suggested to use at least $2K$ collocation points [27]. However, recently it has been shown that when the solution expansion is sparse, not only regression leads to accurate approximations but also the coefficients can be accurately estimated using samples much fewer than the number of coefficients, thereby saving costs in experimentation or high-fidelity simulation. This section follows with the review on the basics of sparse recovery for polynomial chaos expansions.

2.2. Sparse spectral analysis using compressive sampling

Compressive sampling has recently emerged and witnessed growing interest as an efficient and inexpensive signal processing algorithm, particularly applicable in cases where the issue of limited number of samples is present because of slow or expensive sampling procedures [28–32]. When the

signal of interest is sparse, compressive sampling allows signal reconstruction with substantially fewer samples than the Shannon sampling rate, thereby suggesting promising solution for signal recovery when there are a few number of measurements available [33–35].

In the case of spectral stochastic techniques, we seek to fit an analytical model to a sample set obtained from the simulation of a complex high-fidelity model, or from expensive experiments. As a result, the generation of samples is a costly procedure, making compressive sampling an applicable algorithm to these problems. Due to its past successful records in reducing the number of required samples, compressive sampling continues to gain extensive attention in UQ. Particularly, it has been shown that if $u(\Xi)$ is sparse with respect to polynomial bases, the approximation expansion can be constructed using $M \ll K$ random realizations [13, 14, 36–39].

Let us denote random realizations of Ξ by $\{\xi^{(i)}\}$. Also, let $\{u(\xi^{(i)})\}$ and $\{u_k(\xi^{(i)})\}$ respectively denote the evaluations of $u(\Xi)$ and $u_k(\Xi)$ at those realizations. The sample vector $\{u(\xi^{(i)})\}$ can be thought of as the set of empirical samples or the set of simulation results from an exact of high-fidelity model. We define vectors \mathbf{u} and \mathbf{u}_k and matrix Ψ as following:

$$\mathbf{u} := (u(\xi^{(1)}), \dots, u(\xi^{(M)}))^T, \quad (7)$$

$$\mathbf{u}_k := (u_k(\xi^{(1)}), \dots, u_k(\xi^{(M)}))^T, \quad (8)$$

$$\Psi(i, j) := \psi_{\alpha^j}(\xi^{(i)}), \quad i = 1, \dots, M, \quad j = 1 : \dots, K. \quad (9)$$

Matrix Ψ has been referred to as measurement matrix or dictionary in the literature. The above definitions along with Equation 3 simply imply that $\mathbf{u}_k = \Psi \mathbf{c}$, where $\mathbf{c} \in \mathbb{R}^K$ is the vector of PC coefficients c_α . In order for $u_k(\Xi)$ to approximate $u(\Xi)$, the vector \mathbf{c} should satisfy the condition $\|\mathbf{u} - \Psi \mathbf{c}\|_2 \leq \epsilon$, where ϵ is the accuracy tolerance due to truncation error. Since we are interested in the case where $M \ll K$, there exist infinitely many solutions for \mathbf{c} . When \mathbf{c} is sparse and has few significant terms, compressive sampling approach can be borrowed to find the sparsest solution. The sparse recovery approach can be formulated as the following program

$$\min_{\mathbf{c}} \|\mathbf{c}\|_0 \quad \text{subject to} \quad \|\mathbf{u} - \Psi \mathbf{c}\|_2 \leq \epsilon, \quad (10)$$

where $\|\cdot\|_0$ denotes the ℓ_0 -norm, i.e. the number of non-zero components in a vector. This ℓ_0 minimization problem is NP-hard to solve. Therefore, the non-convex discontinuous ℓ_0 -norm is usually replaced with convex continuous ℓ_1 -norm, leading to the linear program

$$\min_{\mathbf{c}} \|\mathbf{c}\|_1 \quad \text{subject to} \quad \|\mathbf{u} - \Psi \mathbf{c}\|_2 \leq \epsilon, \quad (11)$$

or equivalently to the unconstrained linear program

$$\min_{\mathbf{c}} \|\mathbf{u} - \Psi \mathbf{c}\|_2 + \lambda \|\mathbf{c}\|_1, \quad (12)$$

where λ is the regularization parameter. The two constrained and unconstrained formulations can be made equivalent, i.e. for every value of ϵ , there is a value for λ that makes the two programs equivalent. ℓ_1 minimizations of Eq. 11 and 12 are mostly known as the Basis Pursuit Denoising (BPDN) approach [40].

Theoretical results have shown that under certain conditions the solution to the ℓ_1 minimization is indeed the sparsest solution, i.e. it matches the solution to the ℓ_0 minimization [41, 42]. However, it is a non-trivial task to guarantee that these conditions are met for a specific problem. As a result,

it is generally accepted that the ℓ_1 minimization gives suboptimal solutions. However, due to its computationally appealing convex property, it is still most widely used. Efforts have been made to reformulate the problem such that sparser solutions can be obtained. In particular, non-convex penalty functions that further promote sparsity are shown to outperform ℓ_1 -norm. For example, numerical studies have shown that $\ell_{0.5}$ minimization and $\ell_1 - \ell_2$ minimization problems can lead to a sparser solution than that obtained from an ℓ_1 minimization problem [43, 44].

Regardless of the penalty function that is used to promote the sparsity, the first step in the minimization problem is to calculate the measurement matrix Ψ . This requires us to have established the order k for PCE, and have obtained the random input samples $\{\xi^{(i)}\}_{i=1}^M$. In this work, we postulate that in PCE models representing physical systems, among all the possible combinations of dimension and order that approximate $u(\Xi)$ within the inaccuracy tolerance ϵ , there exists a PCE that is the sparsest. For instance, this sparsest solution may be corresponding to a sub-dimensional PCE at a high order. Let us denote the order and dimension of this sparsest PCE by k^* and $d^* \leq d$, respectively. It is critical what to set as the dimension and order of the expansion before the sparse recovery. If our choice of PCE truncation order k is in fact smaller than k^* , we will fail to include sufficient bases in the measurement matrix. As a result, our measurement matrix will not include the bases that the sparsest PCE lives in and thus the compressive sampling solution will not produce the sparsest possible PCE. On the other hand, if the total PCE order k is set larger than k^* , we will have a fatter measurement matrix, with extra bases in addition to those that sufficiently span the sparsest possible PCE. A fat measurement matrix can lead to a large coherence impairing the sparsity of the solution, as will be described in detail later.

We, so far, have briefly described the implication of PCE order on the recovery of the sparsest solution. The choice of dimension will also have implications. Conventionally, the dimension of PCE, d , is the set to be the number of all random inputs to the system. However, in many physical systems, there are input parameters that have a insignificant influence on the output. If these uninfluential parameters are included, it increases the coherence of measurement matrix without significantly improving the chance of accurate recovery, and thus decreases the solution sparsity. Therefore, it is particularly important in compressive sampling, to include in the PCE as few input parameters as needed given the inaccuracy tolerance ϵ . In the next sections, we include the theoretical foundations for these arguments and introduce an incremental algorithm that identifies the influential input parameters by searching for the dimensionality of the sparsest solution, d^* and also identifies the optimal expansion order k^* .

3. Efficient recovery of sparse solution

In an overdetermined case of estimation chaos coefficients, by increasing the order or dimension of the approximation, the accuracy of the approximate PCE is expected to increase. However, in underdetermined cases, when a sparse recovery is pursued given a fixed number of samples, the accuracy does not necessarily increase when a higher order or higher dimensional expansion is used. That is because the mutual coherence of measurement matrix will increase for expansions with higher order or dimension. Coherence or mutual coherence of any matrix $\Psi \in \mathbb{R}^{M \times K}$ is defined as the maximum absolute normalized inner products, i.e., cross-correlation between its columns [13, 45]. Let $\psi_1, \psi_2, \dots, \psi_K \in \mathbb{R}^M$ be the columns of matrix Ψ , the coherence of matrix Ψ , $\mu(\Psi)$, is then given by

$$\mu(\Psi) := \max_{1 \leq i, j \leq K, i \neq j} \frac{|\psi_j^T \psi_i|}{\|\psi_j\|_2 \|\psi_i\|_2}. \quad (13)$$

This section follows with mathematical supports on why a large coherence deteriorates the accuracy of compressive sampling solution. It is worth noting that these results apply to measurement matrices that consist of any bases, except for the last result that is only valid for a measurement matrix consisting of PC bases, hereinafter referred to as PC measurement matrix. For notation brevity, however, we use the notation Ψ interchangeably for both general and PC measurement matrices, while making the distinction in the wording of statements.

3.1. Impact of coherence in sparse recovery

A large body of work in compressive sampling has focused on the accuracy and stability of this approach. In what follows, we provide a number of theorems developed in this field that are related to coherence.

A widely used property in theoretical results in compressive sampling is the restricted isometry property (RIP). Specifically, S -restricted isometry constant of a matrix Ψ is the smallest $\delta_S \in (0, 1)$ such that

$$(1 - \delta_S) \|\mathbf{c}\|_2^2 \leq \|\Psi^{S^*} \mathbf{c}\|_2^2 \leq (1 + \delta_S) \|\mathbf{c}\|_2^2, \quad (14)$$

for every submatrix of $\Psi^{S^*} \in \mathbb{R}^{M \times S^*}$, $S^* \leq S$, of $\Psi \in \mathbb{R}^{M \times K}$ and every vector $\mathbf{c} \in \mathbb{R}^{S^*}$.

Theorem 1 ([46]). *Suppose $\mathbf{u} = \Psi \mathbf{c}^* + \mathbf{z}$ with $\|\mathbf{z}\|_2 \leq \epsilon$, where \mathbf{c}^* is S -sparse ($S > 1$). If $\delta_S < 0.307$, the solution of the ℓ_1 minimization problem given in (11) denoted by $\bar{\mathbf{c}}$ satisfies*

$$\|\mathbf{c}^* - \bar{\mathbf{c}}\|_2 \leq \frac{\epsilon}{0.307 - \delta_S}, \quad (15)$$

or in noiseless situation exactly recovers \mathbf{c}^ .*

Although Theorem 1 does not directly involves coherence, it is closely related to coherence. The following simple proposition explains this connection.

Proposition 1 ([47]). *Suppose that the columns of Ψ are normalized. Then for any $S < \mu(\Psi)^{-1} + 1$, Ψ satisfies $\delta_S \leq (S - 1)\mu(\Psi)$.*

Proposition 1 makes it obvious that a smaller coherence enforces smaller δ_S , thus smaller recovery error. In general, although the RIP-based theorem provides stronger results, it does not offer much applicability. This is because for a given measurement matrix, as opposed to the simplicity in calculating coherence, calculation of RIP constant is known to be an NP-complete problem [47].

The first attempt to apply compressive sampling in the estimation of PC coefficients was reported in [13], where conditions on the stability of estimated coefficients were investigated. Specifically, the following proposition established conditions on the sparsity of PCE solution to ensure stability with a high probability.

Proposition 2 ([13]). *Let $u_k(\Xi)$ be a S -sparse Legendre PC approximation of function $u(\Xi)$ with an index set that is a subset of $\Lambda_{k,d}$. Also, let the \mathbf{c}^* be the vector of PCE coefficients which satisfies the sparsity condition*

$$\|\mathbf{c}^*\|_0 = S < 0.25 \left(1 + \frac{1}{\mu(\Psi)} \right). \quad (16)$$

Then, with probability larger than $1 - \exp\left(-\frac{M}{8(3^{2k})}\right)$, the ℓ_1 minimization solution, $\bar{\mathbf{c}}$, given in (11) obeys

$$\mathbb{E} \left[\left(u_k(\Xi) - \sum_{\alpha \in \Lambda^{k,d}} \bar{c}_\alpha \psi_\alpha(\Xi) \right)^2 \right] \leq \frac{4}{3M} \frac{(\epsilon + \|\Psi \mathbf{c}^* - \mathbf{u}\|_2)^2}{1 - \mu(\Psi)(4\|\mathbf{c}^*\|_0 - 1)}. \quad (17)$$

The impact of coherence is transparent in Proposition 2. The smaller the coherence, the more accurately the ℓ_1 minimization solution can approximate the actual sparse solution.

3.2. Coherence of a measurement matrix consisting of PC bases

The PC measurement matrix, i.e. the measurement matrix built by PC bases, is a particular case of measurement matrices, and thus offers particular property. Specifically, let us consider how changes in the construction of the model impact the coherence of measurement matrices. For instance, if a Fourier series expansion is to be estimated, the model design parameter is the set of frequencies. In this case, if a new richer expansion is to be constructed by adding a single new frequency, the number of columns in measurement matrix will be increased by two. In the case of a PCE model, however, the model design parameters are dimension and order, and a one step increment in each increases the size, and therefore coherence, of the PC measurement matrix, due to Eq. 5. Figure 1 shows numerical results highlighting such changes. Specifically, we have evaluated Legendre polynomials at different combinations of total order and dimension (each varying from 1 to 8), at uniform sample sets of size 100, 200 and 300. It can be seen that the increase in coherence with respect to total order and dimensionality has a higher rate for smaller sample sets.

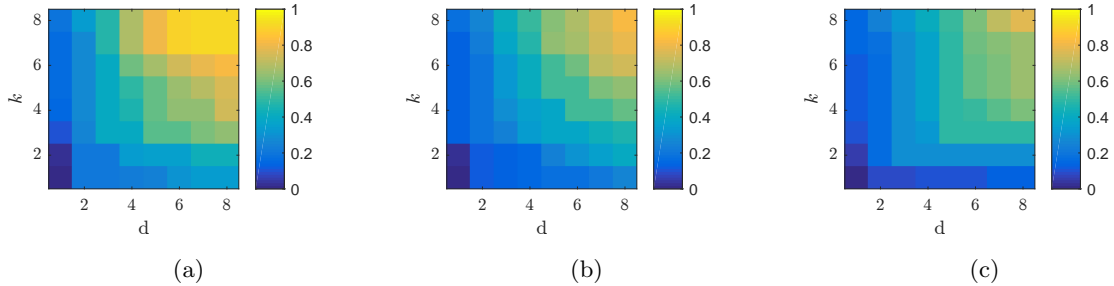


Figure 1: Coherence of PC measurement matrices populated based on evaluations of Legendre polynomials with different combinations of dimension. d and total order, k . The polynomials are evaluated at (a) 100 uniform samples, (b) 200 uniform samples, and (c) 300 uniform samples.

In models for physical systems, it is often the case that not all the input parameters have a significant impact on the output. Therefore, when the number of samples is limited and compressive sampling is to be used, including bases corresponding to non-significant random inputs can undesirably increase the coherence of PC measurement matrix, while failing to provide valuable dictionary for better recovery. This can significantly deteriorate the accuracy of coefficient estimation. It should be noted that Theorem1 holds only when the signal of interest is in fact sparse in the dictionary Ψ . Therefore, a sparse recovery is successful if the particular choice of dictionary can sparsely represent the signal of interest [48].

Recently, efforts have been made in learning the dictionaries, i.e. the best bases, in which signals are sparse, as opposed to using predefined and off-the-shelf dictionaries such as wavelet and cosine dictionaries. Numerical results have shown significant advantages realized by learning the dictionary [48–50]. There are two classes of approaches in dictionary learning: (a) approaches that identifies the best dictionary among a set of candidate dictionaries [51], and (b) more complex approaches that adopt a dictionary from scratch using iterative algorithms [48–50]. In compressive sampling for PCE, since the type of bases is predetermined based on the distribution type of inputs, the first class of approaches is applicable. In what follows, we introduce our incremental algorithm which is able to select the right dictionary given a coherence budget dictated by the number of samples.

3.3. A coherence-aware incremental algorithm for sparse PC approximation

Our objective is to learn, or intelligently choose, the bases in the measurement matrix or dictionary Ψ , given a set of samples. For practical reasons, in working with PCE, we translate this objective into that of identifying the “ideal” combination of dimensionality and expansion order. As suggested in [13] and is apparent from Figure 1, in high dimensional PCEs, the expansion order should be kept very small to avoid large coherence for PC measurement matrix. In such cases, if there really exists a sparser PCE approximation at a higher order, it will not be recovered. Our approach to overcome this builds up on (a) the aforementioned argument that in physical systems, input parameters could rank differently based on their impact on the output, and (b) the possibility that by removing uninfluential parameters from the expansion, a higher order sparser PCE, or in general a sparser PCE, is likely to be recovered. In other words, given the competing influence of dimension and order on coherence, we compromise on the former to allow for the latter to increase without largely affecting the coherence. The success of such sparse recovery hinges on how accurate a lower dimensional expansion can still approximate the system of interest.

To implement this, an incremental selection approach for learning the PC measurement matrix is proposed that identifies the influential input parameters, given the available samples set, and prevents the PCE construction to be unnecessarily high dimensional. We postulate that given a tolerance budget, there exists a sparsest PCE approximation among various combinations of dimension and order levels. To reach this combination, we develop an incremental search whose steps involve one unit trial increases in a dimension or the order. For each trial increment, in dimension or order, a set of polynomial bases at once enter the measurement matrix. To determine which trial should be selected at the end of each step, the following two assumptions furnish the foundation for our approach.

Assumption 1. *If the trial basis set added to the measurement shares bases with those spanning the sparsest PCE approximation, the sparsity of the resulting ℓ_1 minimization solution will increase.*

Assumption 2. *Once all the bases that span the sparsest approximation are included in the measurement matrix, adding new bases will only increase the coherence of measurement matrix and can therefore impair the sparse recovery.*

Assumption 2 is always true, while Assumption 1 may not be true if the additional trial bases increase the coherence of measurement matrix so drastically, that even though informative bases are provided, a sparser solution can still not be recovered. However, this is typically not the case in our incremental algorithm, where trial increments in coherence in each step is usually small.

Based on these two assumptions, the incremental algorithm searching through different dimensions and order levels is developed as follows. At the beginning of step t in this search, let Ξ_t^r

denote a reduced dimension random vector which already includes d_t significant random inputs, i.e. $|\Xi_t^r| = d_t$, $d_t < d$ and $|\cdot|$ loosely denotes the dimension of a random vector. Given this reduced dimension, and an expansion order k_t , the PCE expansion at the beginning of step t is given by

$$u_{k_t, d_t}(\Xi_t^r) := \sum_{\alpha \in \Lambda_{d_t, k_t}} c_\alpha \psi_\alpha(\Xi_t^r). \quad (18)$$

Driven by the two assumptions, the algorithm keeps incrementing the dimension or the order of PCE depending on how the sparsity of ℓ_1 minimization solution is improved, as will be explained in detail. The incremental search algorithm stops when sparsity can no longer be improved. At each step, the algorithm chooses among (a) adding one of the parameters to Ξ_t^r , (b) increasing k_t by one, or (c) termination. As mentioned earlier, the single selection criteria is the sparsity of the solution for each trial increment compared to the sparsity at the previous step.

Specifically, at step t of the search, each one of the remaining random inputs, i.e. those that are not already included in Ξ_t^r , is added to Ξ_t^r once at a time and ℓ_1 minimization is solved and the sparsity of the solution is recorded. An ℓ_1 minimization is also solved for the Ξ_t^r of the previous step with an incremented order k_t . If the sparsity of the solution is not improved in either of these cases, the algorithm stops. Otherwise, if adding the incremented order yields the sparsest solution, then the new sparsest PCE solution at the next step will be $u_{k_t+1, d_t}(\Xi_{t+1}^r)$. If the sparsest solution is caused by the addition of a new input parameter, then that random input is added to Ξ_t^r , and the new sparsest PCE solution will be recorded as $u_{k_t, d_t+1}(\Xi_{t+1}^r)$.

Algorithm 1 shows the pseudocode for our incremental algorithm. The algorithm is initialized with a zero-dimensional 2nd order expansion. Instead of order 2, larger values can also be used depending on the prior knowledge about the true order of the solution. Similarly, if a number of influential inputs are *a priori* known, they can be included in the initial $\Xi_{t=1}^r$.

Algorithm 1 Incremental algorithm for PCE coefficients estimation

- 1: Set $t = 1$, $|\Xi_t^r| = 0$, $k_t = 2$. ▷ Start with a zero-dimensional 2nd order expansion.
 - 2: Set $s^* = 0$ and s_{min} to be a large number ▷ s denotes the number of non-zero coefficients.
 - 3: **while** sparsity is improving ($s^* < s_{min}$) **do**
 - 4: **for** $\{\Xi_i : \Xi_i \notin \Xi_t^r\}$ **do**
 - 5: Temporarily add input parameter Ξ_i , to Ξ_t^r . ▷ $|\Xi_t^r| = |\Xi_t^r| + 1$
 - 6: Set \mathbf{c}_i^* to be the solution of related Eq. 11.
 - 7: Remove input parameter Ξ_i from Ξ_t^r .
 - 8: **end for**
 - 9: Set $s^{(d)}$ to be ℓ_0 -norm of the sparsest recorded solutions \mathbf{c}_i^* .
 - 10: Temporarily add k_t by one, solve the related Eq. 11, and record its solution's ℓ_0 -norm as $s^{(k)}$.
 - 11: $s^* = \min\{s^{(k)}, s^{(d)}\}$ ▷ Pick the sparser solution.
 - 12: **if** $s^* < s_{min}$ **then** ▷ Check if the new solution is sparser than the previous step.
 - 13: Increment the dimension or order based on Line 11.
 - 14: $s_{min} = s^*$.
 - 15: $t = t + 1$.
 - 16: **end if**
 - 17: **end while**
-

At initial steps of the sequence, i.e. for a small t and when d_t and k_t are also small, the number of available samples, M , can be larger than the number of unknown chaos coefficients, $|\Lambda_{d_t, k_t}|$. In such overdetermined cases, to decide whether the increment should be in certain dimension or in the order, instead of inspecting changes in sparsity, we inspect changes in least square error (LSE). It has been shown that minimizing the least square error is an effective approach for basis selection [52]. To decide where to switch the criterion from LSE to sparsity, we make use of the desired accuracy value, ϵ in Eq. 11, and while LSE is larger than this desired value, instead of solving ℓ_1 minimization and recording the sparsity, we solve the following minimization

$$\min_{\mathbf{c}} \|\mathbf{u} - \Psi \mathbf{c}\|_2^2. \quad (19)$$

3.4. Note on computational cost

The proposed incremental approach involves solving Eq. 19 and Eq. 11 multiple times for PCEs with different dimension and order. This, however, does not result in a substantial computational cost as these equations can be solved with very efficient algorithms. At initial steps, where ℓ_2 minimization problem in Eq. 19 is solved, the computational cost is of $\mathcal{O}(MK^2)$, mainly due to matrix-matrix multiplication. It should be noted that initial steps are applied to low dimensional low order expansions where K is small, and in cases which does not involve large number of samples, the algorithm soon switches to ℓ_1 minimization of Eq. 11. To solve this ℓ_1 minimization problem, BPDN solver in SPGL1 package [53], can be used, which runs a fast algorithm with computational cost that is mainly due to matrix-vector multiplication with complexity of $\mathcal{O}(MK)$ [54].

The final cost of incremental algorithm depends on the number of steps before reaching the sparsest solution, which depends on d^* and k^* . Therefore, the additional cost incurred by this incremental approach will be moderate when only a small subset of random parameters are in fact influential. As an example, in an 80-dimensional problem, which is described in detail in Section 4.3, where $d^* = 5$, $k^* = 3$, and $M = 500$, the incremental algorithm took only 50.56 sec on a personal computer. It should be noted that in achieving the final goal of constructing a PCE to approximate a complex model, this additional cost is typically negligible compared to the large sampling cost imposed by expensive simulations of the model under study. Also, the additional cost should be considered in conjunction with the significant benefits materialized as a result of a more accurate approximation and a lower dimensional representation.

In next section, we provide the results of numerical validation for the proposed algorithm. Hereafter, we refer our algorithm as *incremental*, and refer to ℓ_1 minimization with a full expansion as *conventional*.

4. Numerical examples

In this section we provide four different numerical examples to demonstrate the validity of our proposed algorithm and compare it with other approaches. Uniformly distributed random variables together with Legendre polynomials are used in all examples. In the first example, the target function is a manufactured sparse low dimensional polynomial expansion with arbitrary coefficients. In the second example, the solution of a stochastic diffusion problem is considered. The third example concerns approximation of a sparse high dimensional polynomial expansion. Finally, in the fourth example, we address approximating community land model using our proposed algorithm. The first three examples are designed in a way that a subset of input parameters are influential. The forth example, on the other hand, is a high dimensional real world problem where

the number of influential random input is not known. We have used the MATLAB package `SPGL1` [53] to solve the ℓ_1 minimization problem in 11 in the conventional approach and also at each step in our incremental algorithm. We have used the MATLAB function `lsqlin` for LSE minimization in the initial steps of the incremental algorithm. In all examples, we start the incremental algorithm with $k_1 = 2$ and set Ξ_1^r to be an empty vector.

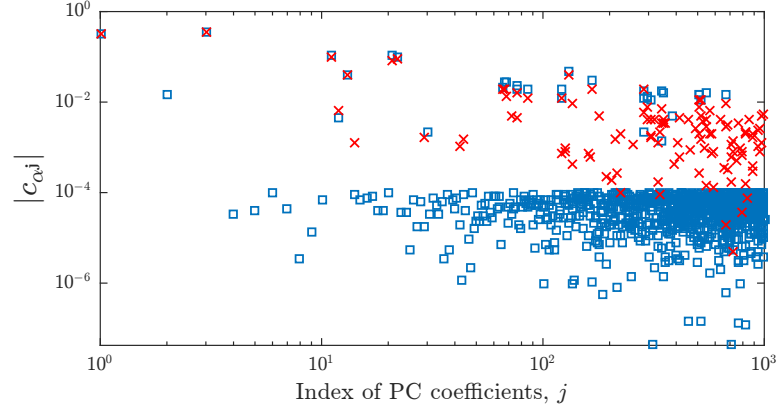
4.1. Example 1: A low dimensional manufactured polynomial expansion

In this example, we manufacture a target function as a 4th order Legendre polynomial expansion in a 10-dimension random input with uniform density on $[-1, 1]^{10}$. This example is designed to evaluate the performance of different algorithms when the output PCE in fact lives in a smaller dimension than that of the input parameter space. Accordingly, we specify values for the 1001 chaos coefficients such that the target polynomial expansion has significant coefficients in 3 dimensions. Specifically, for each coefficient c_α , if it only includes any of the first three random inputs, we assign a random realization from $U(-e^{-\|\alpha\|_1}, e^{-\|\alpha\|_1})$. Otherwise, we assign to all remaining coefficients a random realization from $U(-10^{-4}, 10^{-4})$. A visualization of the coefficient values is shown as squares in Figure 2.

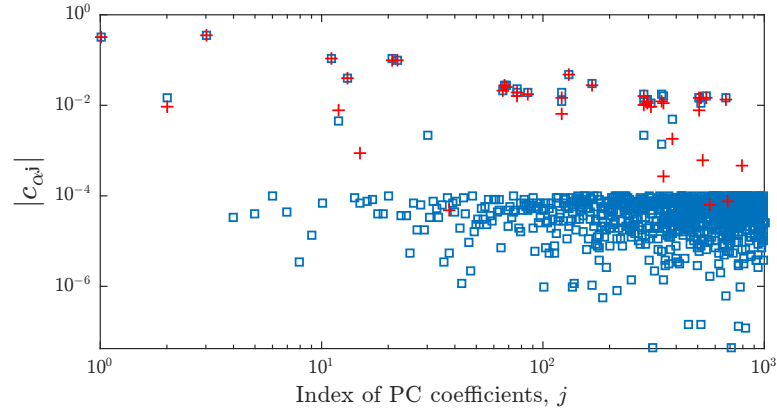
Our task is to evaluate the performance of our incremental and also the conventional algorithm in recovering the expansion coefficients, if a small number of samples from this manufactured expansion are provided. We set $\epsilon = 0.01 \|\mathbf{u}\|_2$. For conventional approach, we consider two cases: (a) when there is complete ignorance about the influential inputs and the measurement matrix is evaluated for $d = 10$, and (b) when out of the seven insignificant inputs, five are known to the modeler and as a result the measurement matrix is evaluated for $d = 5$. The choice of order depends on the number of available samples and also the dimensionality of the expansion. The order must be chosen large enough to ensure the desired accuracy, but nonetheless small enough to prevent the problem from being *heavily* under-sampled and the measurement matrix from being highly coherent. However, there exists no rigorous criterion to determine *a priori* when a problem is *heavily* under-sampled, or in other words, what is the relationship between the number of samples and the largest k that can be used. Prior knowledge on the smallest order that provides the desired accuracy doesn't exist, either. Therefore, the choice of order is typically ad hoc. In this example, for the 10-dimensional case we set $k = 4$, while for the 5-dimensional case, which may allow for a higher order, we set $k = 5$.

Figure 2 compares the true coefficients with the recovered ones for a single trial using 100 samples. It can be seen that conventional algorithm results in a less sparse solution. The larger number of non-zero coefficients, in this case, is the direct result of large coherence of the PC measurement matrix populated in a conventional way with high dimension and order. Namely, when the columns in the measurement matrix are closely related, it becomes less easy to distinguish which basis in fact impact the output [55]; hence a convoluted attribution of impacts.

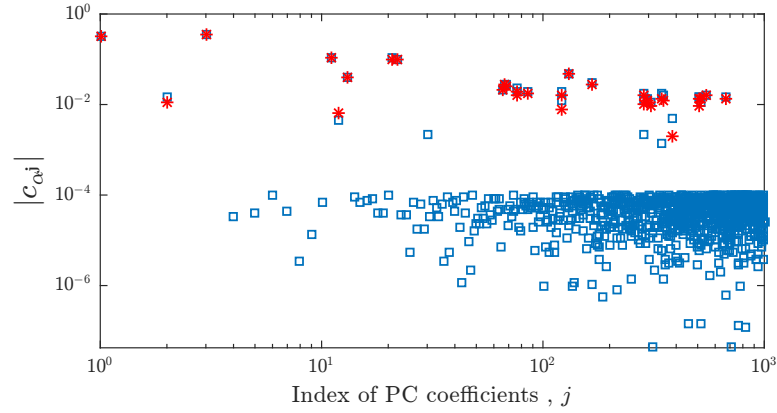
Figure 3a compares the success rate of coefficient recovery for incremental and conventional approaches. A successful recovery in this example is defined to be one that satisfies $\|\bar{\mathbf{c}} - \mathbf{c}^*\|_2 / \|\mathbf{c}^*\|_2 \leq 0.02$, where $\bar{\mathbf{c}}$ and \mathbf{c}^* are the vectors of recovered and exact coefficients, respectively. Figure 3b compares relative validation error, defined as $\|\bar{\mathbf{u}} - \mathbf{u}\|_2 / \|\mathbf{u}\|_2$, where \mathbf{u} includes evaluations of target (exact) expansion at 200 new samples from the inputs and $\bar{\mathbf{u}}$ includes the evaluations of recovered expansion at the same 200 sampled inputs. Figure 3c compares the number of non-zero terms in recovered solutions, and finally 3d compares the dimension of recovered polynomials. In these figures, for a more robust comparison, for each sample size M , 100 independent sample sets are generated, based on which the shown values are average quantities over 100 recovery trials. It can



(a) Conventional approach with $d = 10$ and $k = 4$



(b) Conventional approach with $d = 5$ and $k = 5$



(c) Incremental approach

Figure 2: Comparison between true and recovered coefficients for the 4th-order 10-dimensional manufactured polynomial expansion of Section 4.1. The target function is manufactured so that it can be effectively approximated using a 4th order 3-dimensional expansion. True coefficients are shown by squares.

be seen that incremental algorithm significantly outperforms conventional algorithm by uncovering of a lower dimensional structure. Besides, Figure 3 also suggests that when the predetermined dimension and order in conventional approach are set closer to the “significant” dimension and order, better results can be obtained.

4.2. Example 2: A stochastic partial differential equation

In this example, we consider the solution of the following stochastic linear elliptic partial differential equation (PDE)

$$\begin{aligned}\nabla \cdot (a(\mathbf{x}, \Xi) \nabla u(\mathbf{x}, \Xi)) &= 1, \quad \mathbf{x} \in \mathcal{D}, \\ u(\mathbf{x}, \Xi) &= 0, \quad \mathbf{x} \in \partial\mathcal{D},\end{aligned}\tag{20}$$

where $\mathcal{D} = (0, 1) \times (0, 1)$ with boundary $\partial\mathcal{D}$, and a is the random diffusion coefficient given by

$$a(\mathbf{x}, \Xi) = a_0 + \sigma \sum_{i=1}^d \sqrt{\lambda_i} \phi_i(\mathbf{x}) \Xi_i,\tag{21}$$

where $a_0 = 1$, $\sigma = 0.3$, $d = 20$, $\{\Xi_i\}_{i=1}^d$, are uniform random variables each taking values in $[-1, 1]$, and $\{\lambda_i\}_{i=1}^d$ and $\{\phi_i(\mathbf{x})\}_{i=1}^d$ are the d largest eigenvalues and eigenfunctions of the exponential kernel

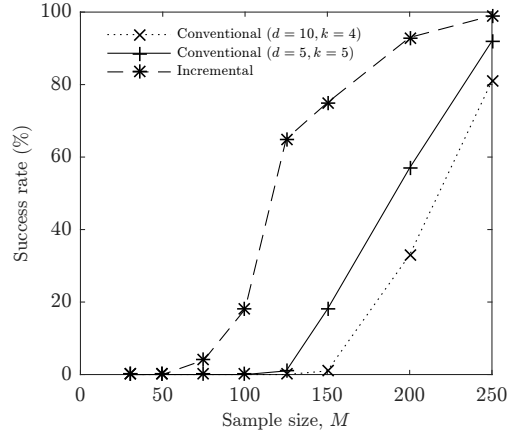
$$C(\mathbf{x}, \mathbf{y}) = \exp \left[\frac{-|x_1 - y_1| - |x_2 - y_2|}{l} \right],\tag{22}$$

with correlation length $l = 1$.

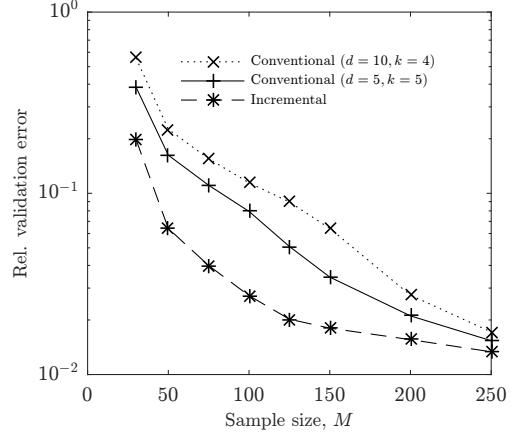
For each sample of Ξ , we evaluate $u((0.61, 0.58), \Xi)$ in Eq. 20 using MATLAB’s `pdenonlin` as the quantity of interest (QoI). We use our incremental approach together with the conventional compressive sampling approach with $\epsilon = 0.002 \|\mathbf{u}\|_2$ to obtain PCE approximations of QoI. In this example, since we do not prior knowledge about d^* and k^* , we include all random inputs in evaluation of measurement matrix for conventional approach. Also, since d is large, we need to choose a small expansion order k . For a better comparison between conventional and our incremental algorithms, we run the conventional algorithm for two different k values, $k = 2$ and $k = 3$. Figure 4 compares the average results for 100 independent trials for incremental and conventional algorithms. Figure 4a compares the success rates. We consider $\bar{\mathbf{c}}$ to be a successful approximation of PCE coefficients if it satisfies $\|\Psi \bar{\mathbf{c}} - \mathbf{u}\|_2 / \|\mathbf{u}\|_2 \leq 0.003$, where \mathbf{u} is a vector of evaluation of target function at 200 new “test” samples. Figure 4b compares relative validation errors for the 200 test samples. Figure 4c compares the number of non-zero terms in recovered solutions, and finally 4d compares the dimension of solution PCE after the sparse recovery. As is evident by the results, the incremental algorithm outperforms the conventional algorithm in terms of sparsity and accuracy by exploring the reduced dimension PCE approximation of target function. It can be seen that these improvements in the results are more significant when the sample size is small and as a result coherence is relatively larger for the measurement matrix with prefixed k and d . In this example, k^* is found by the incremental approach to be 2. We observe that the closer the prefixed p and d are to the p^* and d^* , the better conventional algorithm can recover the sparsest solution. However, this may not be the case all the time, and therefore the incremental algorithm should be preferred.

4.3. Example 3: A high dimensional manufactured polynomial expansion

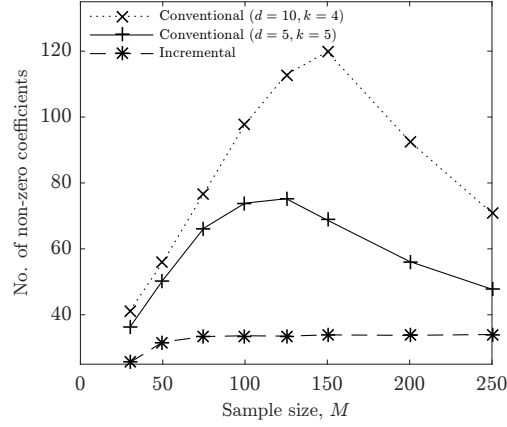
In the previous two examples, the prefixed expansion order, k , in conventional approach was equal or greater than the order of the underlying solution, k^* . Therefore, although the measurement



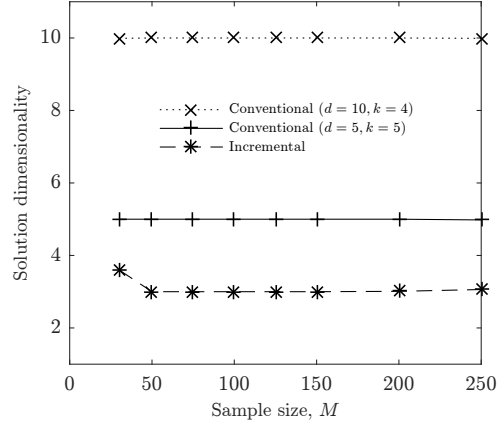
(a)



(b)

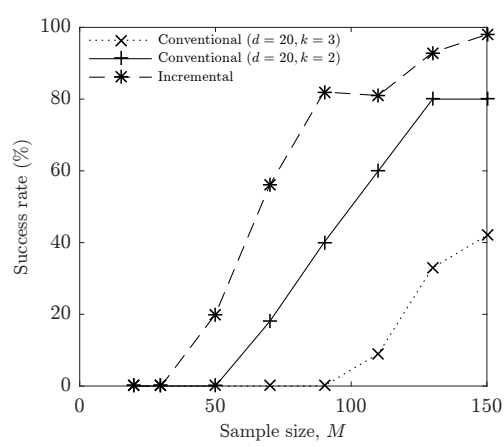


(c)

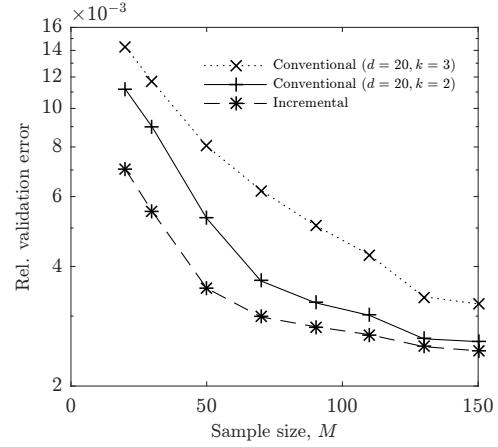


(d)

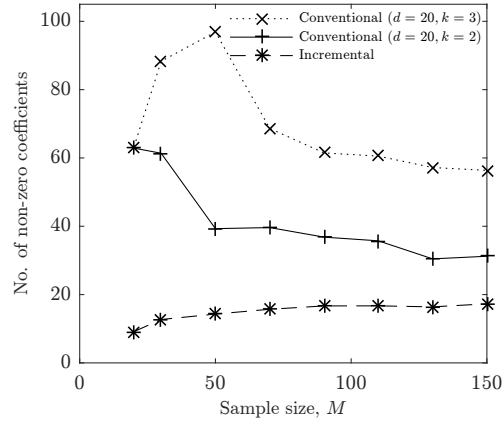
Figure 3: Comparison of different performance metrics for incremental and conventional algorithms for the 4th-order 10-dimensional manufactured polynomial expansion of Section 4.1. Markers show values for (a) success rate (b) relative validation error (c) number of non-zero coefficients (d) solution dimensionality, averaged over 100 independent trials for each sample size.



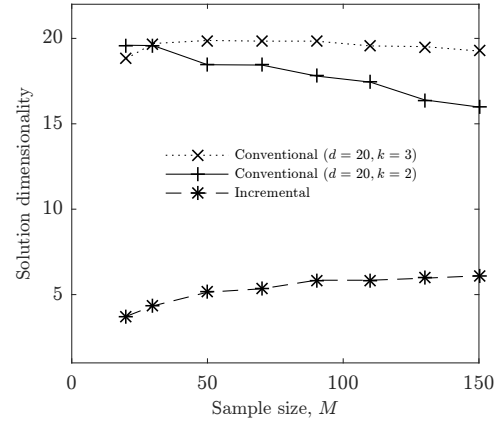
(a)



(b)



(c)



(d)

Figure 4: Comparison of incremental and conventional algorithms results for stochastic PDE (a) relative validation error (b) number of non-zero coefficients (c) solution dimensionality.

matrix in conventional approach suffered from large coherence, it did contain all the bases that the PCE approximation was sparse in. In other words, the conventional approach was ineffective only because it could not identify the few important bases out of a wide set of choices. In this example, we consider a different class of problems where conventional approach can be ineffective. Specifically, we consider problems whose high dimensionality force the prefixed expansion order k in conventional approach to be set to a small value, likely to be lower than the underlying solution order k^* . As a result, the measurement matrix in conventional approach will lack important higher order bases. Therefore, the conventional approach not only suffers from a large coherence and poor sparse recovery, but also from the inability to form a sufficiently wide set of candidate bases to be identified through the recovery process.

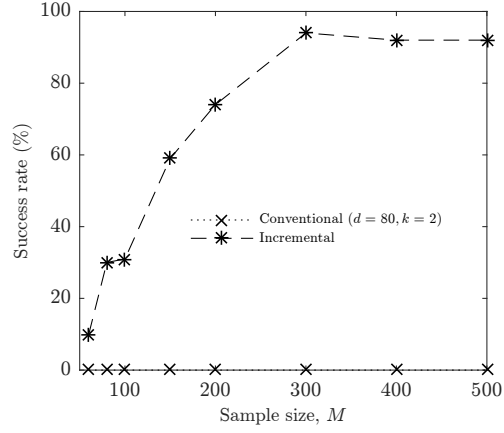
To demonstrate the advantage of the incremental approach, let us consider the target function, $u(\Xi)$, to be a third order 80-dimensional polynomial expansion given by

$$u(\Xi) = \sum_{i=1}^{K_1} \Xi_i + \sum_{i=1}^{K_1-1} \Xi_i \Xi_{i+1} + \sum_{i=1}^{K_1-2} \Xi_i \Xi_{i+1} \Xi_{i+2} + \sum_{i=K_1+1}^K \frac{1}{5(1+i)^2} \Xi_i + \sum_{i=K_1+1}^{K-1} \frac{1}{3(1+i)^2} \Xi_i \Xi_{i+1}, \quad (23)$$

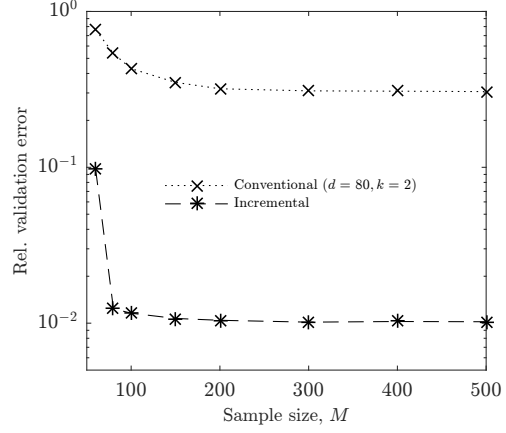
where $K_1 = 5$ and $K = 80$ and Ξ is the vector of uniformly distributed random inputs with values in $[-1, 1]^{80}$. Evaluations of this target function at sample points in the parameter space are used as observation samples in the sparse recovery algorithms. For the conventional approach, we set $d = 80$ and $k = 2$, and note that a third order expansion would lead to a measurement matrix with 91,881 basis functions and the problems becomes drastically under-sampled. Also, we set $\epsilon = 0.01 \|\mathbf{u}\|_2$. As expected, the incremental algorithm results in $k^* = 3$. Figures 5 compares the average results for 100 independent trials for incremental and conventional algorithms. Figure 5a compares the success rates. We consider $\bar{\mathbf{c}}$ to be a successful approximation of PCE coefficients if it satisfies $\|\Psi \bar{\mathbf{c}} - \mathbf{u}\|_2 / \|\mathbf{u}\|_2 \leq 0.011$, where \mathbf{u} is a vector of evaluations of target function at 200 new samples. It can be seen that the success rate remains at zero for conventional approach as it never achieves the desired accuracy. Also, in Figure 5b, where the relative validation errors for 200 independent test samples are shown, it can be seen that conventional approach results in very large prediction error. This is a direct result of using a measurement matrix that does not include important higher order bases. Furthermore, it can be seen in Figures 5c and 5d, that both the number of non-zero terms and dimensionality of the conventional solution increases as more samples become available. The reason is that it gets more challenging to meet the desired ϵ with more samples considering that the measurement matrix lacks some essential bases.

4.4. Example 4: A community land model

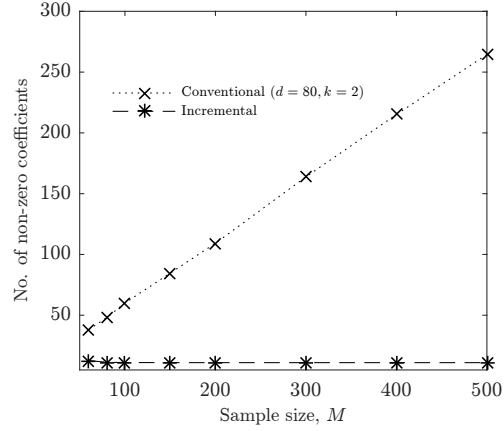
Previous examples were designed such that only a small subset of random inputs have a significant impact on output. In this example, we consider a real world problem where the number of significant random inputs is not *a priori* known. We aim to develop a PC surrogate for a computationally expensive community land model (CLM) using our incremental algorithm. CLM is the land component of community earth system model (CESM). CLM is used to predict the future states of climate and ecosystems including the future condition of carbon cycles and vegetation dynamics. In this paper, we consider the community land model with carbon/nitrogen biogeochemistry (CLM-CN). In [56], offline CLM-CN simulations were performed to predict the steady-state vegetation for Niwot Ridge flux tower site in Colorado for a 100 year time horizon. Five different output quantities of interest were considered: leaf area index, total vegetation carbon, gross primary production, heterotrophic respiration and photosynthesis. These models are computationally expensive, with



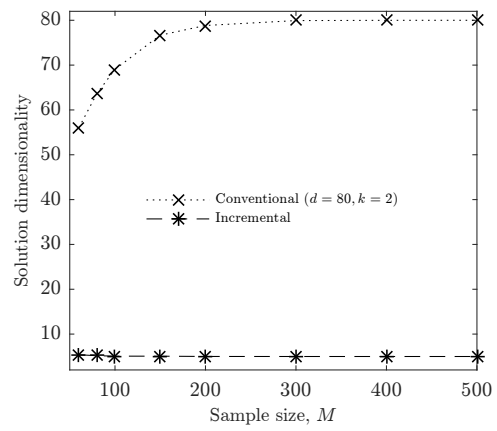
(a)



(b)



(c)



(d)

Figure 5: Comparison of incremental and conventional algorithms results for manufactured high dimensional PCE (a) relative validation error (b) number of non-zero coefficients (c) solution dimensionality.

the computation time of a single simulation reaching 10 hours on a single processor [56]. Thus, a surrogate for this expensive model can significantly reduce the simulation cost and allow further analysis. In this paper, we seek to develop a surrogate model for the calculation of leaf area index (u_{LAI}), as a scalar QoI, using a PC expansion over the input parameters. Out of the 81 uniformly distributed CLM input parameters listed in [56], 2 parameters can be written as functions of the other parameters. Therefore, only 79 parameters are independent and are used as random inputs in the Legendre-based PC expansion.

In [56], when the QoI, u_{LAI} , is smaller than 0.3, it was assumed to that the system is in a dead vegetation state. The probability space, Ω , is then considered to consist of two mutually exclusive subspaces; one that results in dead vegetation state, and one with $u_{LAI} \geq 0.3$, which is denoted by \mathcal{A} . The reason that $u_{LAI} \leq 0.3$ is considered as dead vegetation is that it is believed that the simulations resulting in small u_{LAI} would have converged to zero if simulations were run for a longer time. In this work, we seek to approximate the output over the latter subspace, by building a PCE for the live vegetation state. We do so by only using the samples with $u_{LAI} \geq 0.3$. Specifically, the sample set includes 9983 simulation results, out of which 3437 samples belong to live vegetation state [56].

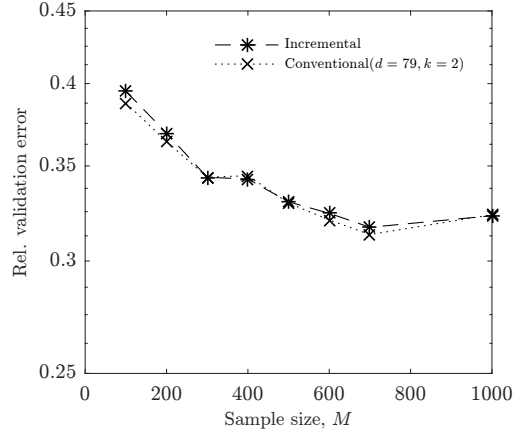
Since u_{LAI} is by definition takes non-negative values, the logarithm of u_{LAI} is approximated with a PCE, given by

$$u_{LAI}(\Xi) = \begin{cases} \exp(\sum_{\alpha} c_{\alpha} \psi_{\alpha}(\Xi)), & \text{if } \Xi \in \mathcal{A}, \\ 0, & \text{otherwise.} \end{cases}$$

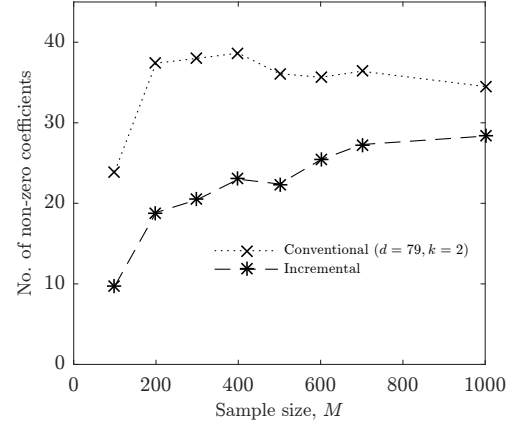
We estimate the PC coefficients using both conventional and incremental algorithms. For conventional approach, we include all 79 random inputs and evaluate the measurement matrix for $k = 2$. In this example, instead of prefixing the inaccuracy tolerance ϵ , we use cross-validation approach to select ϵ , to avoid overfitting. Cross-validation results showed that $\epsilon = 0.3 \|\mathbf{u}\|_2$ results in the lowest prediction errors computed using validation samples. Although this accuracy level might seem too low, it is in fact reasonable considering the high dimensionality of the problem and the limited number of samples.

Figure 6 shows the average results for 20 independent trials for incremental and conventional algorithms. Figure 6a compares the average relative validation error for 200 test samples. Although incremental algorithm does not result in significant reduction in prediction error, i.e., most probably because of very small sample sizes, as it can be seen in Figure 6b and 6c, it considerably reduces the dimensionality and sparsity of the approximation solution. Invoking Occam's razor principle, such lower dimensional models may be preferred over higher dimensional ones [57].

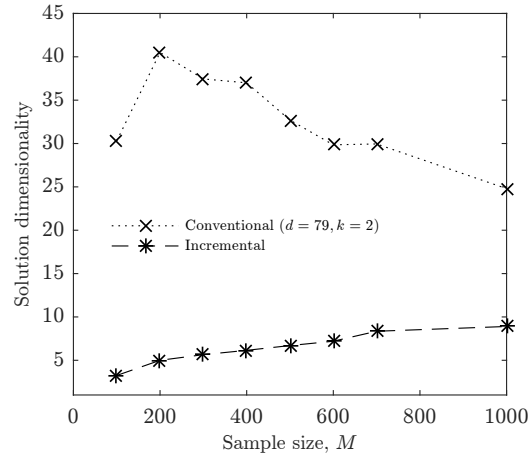
To evaluate how successful our algorithm is in dimension reduction and identifying the right significant parameters, we compare our results with the variance-based sensitivity analysis performed in [56] using the PC surrogate estimated by their proposed Bayesian compressive sensing (BCS). To do so, we first define an importance metric quantified for each random input and averaged over independent trials. A number of independent trials are considered merely in order to make the results less sensitive to a particular sample set, as the problem is severely undersampled. Our choice of importance metric considers the number of times an input parameter, i.e. a particular dimension, appears in the identified sparsest solution across all trials. It also notes the order in which parameters are added to optimal dimension during the incremental algorithm, i.e. its order in the final dimension reduced input vector Ξ^r . Specifically, for each random input, Ξ_i , the overall



(a)



(b)



(c)

Figure 6: Comparison of incremental algorithm and conventional BPDN results for community land model (a) relative validation error (b) number of non-zero coefficients (c) solution dimensionality.

importance metric, I_i , is defined as

$$I_i := \sum_{j=1}^{N_{\text{trials}}} \frac{(d - a_{ij})\delta_{ij}}{d}. \quad (24)$$

If dimension i is identified in the sparsest solution in trial j , then $a_{i,j}$ denotes its order in the identified input vector, and $\delta_{ij} = 1$; otherwise, $a_{i,j} = 0$, $\delta_{ij} = 0$. In Figure 7, square markers show the importance metrics quantified for all random inputs based on trials for the largest sample sizes, $M = 1000$, i.e, $N_{\text{trials}} = 20$. We consider a parameter to be uninfluential if it appears in the solution of incremental algorithm fewer than 3 times in the 20 trails. It can be seen that there is a clear separation between importance metrics for eleven of input parameters and that for the remaining 68 uninfluential parameters.

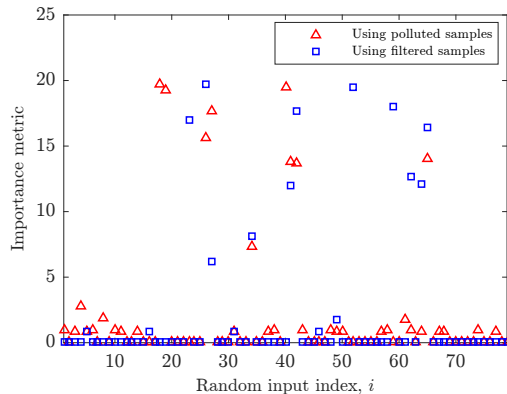


Figure 7: Importance metrics calculated for all random inputs according to Eq. 24. Filtered samples include only the realizations that lead to non-dead vegetation state. Polluted samples consist of filtered samples (85%) and realizations leading to a dead vegetation state (15%).

The eleven influential random inputs are shown on the first row of Table 1. For comparison, the ranking by the variance-based sensitivity/BCS approach reported in [56] is also included in the second row of Table 1. It seems that our algorithm does not identify the top three influential parameters that are reported in [56], i.e. leaf_cn, rc_npool and flnr. This discrepancy can be explained by noting that in the BCS approach, realizations that lead to the dead vegetation state are classified and then filtered out based on an algorithm with 15% likelihood of misclassification. In order to make a consistent comparison between the two rankings, we simulate their filtering process by keeping 15% of realizations with $u_{LAI} \leq 0.3$, and removing 15% of realizations with $u_{LAI} > 0.3$ to form a “polluted” dataset, out of which M random samples are drawn. We still considered 20 trials with $M = 1000$. Based on the newly calculated importance metric, only nine parameters are found to have a large importance metric. The third row in Table 1 shows the ranking of these nine random inputs. We observe that the nine influential random inputs obtained by our algorithm are the same as nine most influential random inputs recognized in [56], however in different order. As an alternative approach, the parameters are also ranked by analytically calculating, for the PCE identified by our incremental algorithm, the Sobol’ indices according to [58, 59]. This ranking is

shown in the last row of Table 1, and as it can be seen, is the same as the importance metric I ranking, except for the order of three parameters that are slightly different. In fact, it was observed that the significance of these three parameters are almost equal, both based on their respective Sobol' indices and also based on the importance metric shown in Figure 7 (the three parameters are indexed at 65, 41, and 42. We conclude that our algorithm is capable of identifying influential random inputs and can be used for dimension reduction.

Table 1: Comparison between the rankings of influential parameters in the community land model identified by the Bayesian compressive sampling approach proposed in [56] and the incremental ℓ_1 minimization (using filtered and polluted samples). The parameters in the shaded cells are identified as uninfluential.

| Algorithm | Ranking of random inputs (left to right) | | | | | | | | | | |
|---|--|----------|------------|-----------|------------|--------|-----------|--------|-----------|-----------|-----------|
| Incremental/metric I filtered data | foot_leaf | rf.s3s4 | k.s4 | q10_mr | frootcn | r_mort | dnp | q10_hr | br_mr | leaf_long | stem_leaf |
| BCS approach of [56] | leafcn | rc_npool | froot_leaf | leaf_long | flnr | r_mort | stem_leaf | br_mr | q10_mr | rf.s3s4 | dnp |
| Incremental/metric I polluted data | leafcn | rc_npool | flnr | stem_leaf | froot_leaf | r_mort | br_mr | q10_mr | leaf_long | qe25 | rhosnir |
| Incremental/Sobol' polluted data | leafcn | rc_npool | flnr | stem_leaf | froot_leaf | br_mr | q10_mr | r_mort | leaf_long | qe25 | rhosnir |

4.5. Other penalty functions

The proposed incremental approach is not limited to ℓ_1 minimization problems, and can improve the performance of other variates of sparse recovery algorithms that use penalty functions other than ℓ_1 -norm, such as weighted ℓ_1 minimization [60], ℓ_p minimization [43], and $\ell_1 - \ell_2$ minimization algorithms [44]. The reason is that if the PC measurement matrix does not contain the bases in which the PCE is sparsest in, regardless of what penalty function is used, the solution's accuracy is heavily impaired. Also, in all compressive sample approaches, a highly coherent measurement matrix always leads to a lower solution sparsity and accuracy, regardless of the type of sparsity promoting penalty function. To demonstrate this in an example, let us consider the $\ell_1 - \ell_2$ minimization algorithm. Numerical results have shown that using $\ell_1 - \ell_2$ -norm instead of the widely used ℓ_1 -norm can lead to better performance [44]. Theorem 2 establishes bounds on $\ell_1 - \ell_2$ recoverability.

Theorem 2 ([44]). *Let \mathbf{c} be an S -sparse vector satisfying*

$$a(S) := \left(\frac{\sqrt{3S} - 1}{\sqrt{S} + 1} \right)^2 > 1. \quad (25)$$

Also, let $\mathbf{u} = \Psi \mathbf{c}^ + \mathbf{z}$ with $\|\mathbf{z}\|_2 < \epsilon$, and suppose Ψ satisfies the condition*

$$\delta_{3S} + a(S)\delta_{4S} < a(S) - 1. \quad (26)$$

Then $\bar{\mathbf{c}}$, the solution of $\ell_1 - \ell_2$ minimization problem

$$\min_{\mathbf{c}} \|\mathbf{c}\|_1 - \|\mathbf{c}\|_2 \quad \text{subject to} \quad \|\mathbf{u} - \Psi \mathbf{c}\|_2 \leq \epsilon, \quad (27)$$

satisfies

$$\|\mathbf{c}^* - \bar{\mathbf{c}}\|_2 \leq \frac{2\epsilon\sqrt{1+a(S)}}{\sqrt{a(S)(1-\delta_{4S})} - \sqrt{1+\delta_{3S}}}. \quad (28)$$

According Proposition 1, smaller coherence for Ψ enforces smaller δ_{3S} and δ_{4S} and results in smaller bound on the recovery error, and vice versa. Therefore, the issue of coherence is present in $\ell_1 - \ell_2$ minimization algorithm, as well, and as such our approach is expected to improve the results.

To numerically demonstrate this, let us consider the target function used in Section 4.1, and the sparse recovery of the associated PC approximations using ℓ_1 and $\ell_1 - \ell_2$ minimization algorithms. We make use of the algorithm developed in [44] for $\ell_1 - \ell_2$ minimization, which solves the following unconstrained minimization

$$\min_{\mathbf{c}} \|\mathbf{u} - \Psi \mathbf{c}\|_2^2 + \lambda(\|\mathbf{c}\|_1 - \|\mathbf{c}\|_2). \quad (29)$$

To make this unconstrained minimization equivalent to the constrained minimization given in (27), we used an iterative approach to calculate the corresponding λ that matches the prescribed ϵ . Figure 8 compares the results for conventional and incremental $\ell_1 - \ell_2$ minimization. It can be seen that regardless of the penalty function, incremental algorithm significantly outperforms the conventional algorithm.

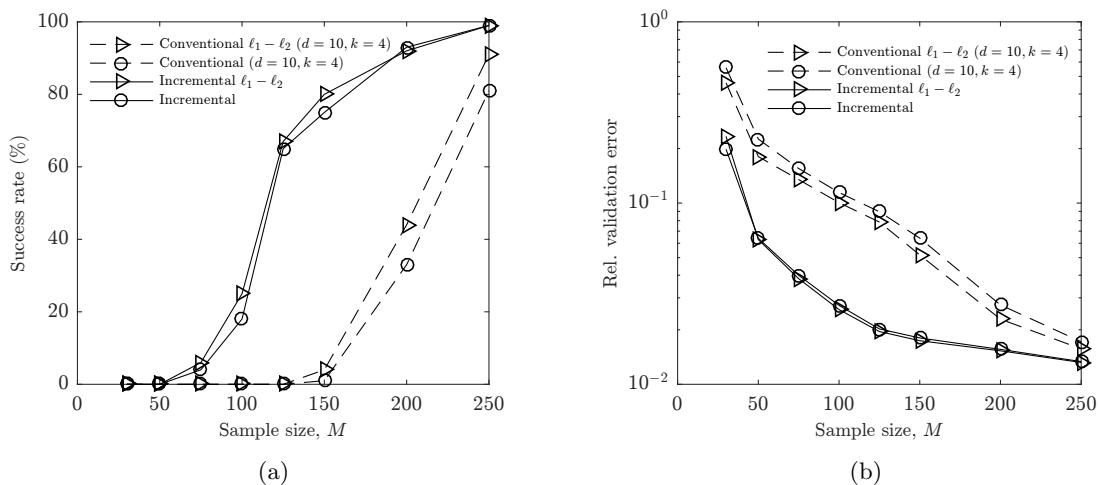


Figure 8: Comparison of incremental and conventional $\ell_1 - \ell_2$ minimization results for manufactured polynomial expansion (a) success rate (b) relative validation error.

5. Conclusion

In this paper, we proposed an incremental compressive sampling algorithm for sparse approximation of PC expansions, when limited number of samples are available. The proposed algorithm seeks to identify the sparsest PC approximation, among expansions with various dimensions and orders. We demonstrated the effectiveness of these algorithms in problems where the solution lies in a lower dimensional subspace in the parameter space. The proposed approach resolves the challenge posed by highly coherent measurement matrices, relevant in high dimensional expansions, by first constituting a very low dimensional candidate solution, and then incrementally adding its dimension or order, depending on the improvement in sparsity. Numerical examples are provided to validate the proposed incremental algorithm. The results show that a substantial improvement

in solution sparsity and accuracy can be achieved by using the incremental algorithm instead of conventional compressive sampling approach.

Our algorithm is different from the adaptive algorithm proposed in [22], in terms of the criterion for selecting new bases or stopping the algorithm, and also the choice of initial basis set. Another important distinction is that if the user has a specific preference for the inaccuracy tolerance ϵ , our algorithm will provide the associated sparsest solution. In other words, our incremental algorithm can perform a tolerance-dependent sparse recovery or dimensionality reduction. Specific tolerance levels can, for instance, be rigidly prescribed by a user based on expert judgment or knowledge about the measurement noise, instead of a cross-validation procedure.

The proposed algorithm can be further refined to become more efficient by an anisotropic trial basis expansion. Specifically, one can refine how expansion order is increased in each trial in the algorithm. Currently, at a dimensionality level, a trial order increase is done by incrementing the total order by one and isotropically impacting all the dimensions. A more refined anisotropic basis set expansion can involve at each iteration multiple one-dimensional order increment trials, each increasing the order in one of the dimensions by one. This can prevent undesirably large increases in the coherence after a one-step degree increment in the algorithm. For problems with a small underlying dimensionality, this gain may not be substantial and the current form of the algorithm is expected to perform well. In cases with larger underlying dimensionality, however, such refinement can potentially lead to a more precise identification of the sparsest solution.

Acknowledgement

The authors would like to acknowledge Dr. Khachik Sargsyan for providing the Community Land Model simulation data which was used in the numerical example of Section 4.4, and is described in detail in [56].

References

References

- [1] J. S. Liu, Monte Carlo strategies in scientific computing, Springer Science & Business Media, 2008.
- [2] R. E. Caflisch, Monte Carlo and quasi-Monte Carlo methods, *Acta numerica* 7 (1998) 1–49.
- [3] S. Heinrich, Multilevel monte carlo methods, in: *International Conference on Large-Scale Scientific Computing*, Springer, 2001, pp. 58–67.
- [4] R. G. Ghanem, P. D. Spanos, *Stochastic finite elements: a spectral approach*, Courier Corporation, 2003.
- [5] D. Xiu, *Numerical methods for stochastic computations: a spectral method approach*, Princeton University Press, 2010.
- [6] E. Novak, K. Ritter, The curse of dimension and a universal method for numerical integration, in: *Multivariate approximation and splines*, Springer, 1997, pp. 177–187.
- [7] V. Barthelmann, E. Novak, K. Ritter, High dimensional polynomial interpolation on sparse grids, *Advances in Computational Mathematics* 12 (4) (2000) 273–288.

- [8] B. Ganapathysubramanian, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems, *Journal of Computational Physics* 225 (1) (2007) 652–685.
- [9] M. Hegland, Adaptive sparse grids, *Anziam Journal* 44 (2003) 335–353.
- [10] T. Gerstner, M. Griebel, Dimension-adaptive tensor-product quadrature, *Computing* 71 (1) (2003) 65–87.
- [11] J. Foo, G. E. Karniadakis, Multi-element probabilistic collocation method in high dimensions, *Journal of Computational Physics* 229 (5) (2010) 1536–1557.
- [12] X. Ma, N. Zabaras, An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations, *Journal of Computational Physics* 229 (10) (2010) 3884–3915.
- [13] A. Doostan, H. Owhadi, A non-adapted sparse approximation of PDEs with stochastic inputs, *Journal of Computational Physics* 230 (8) (2011) 3015–3034.
- [14] L. Mathelin, K. Gallivan, A compressed sensing approach for partial differential equations with random input data, *Communications in computational physics* 12 (04) (2012) 919–954.
- [15] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of Computational Physics* 230 (6) (2011) 2345–2367.
- [16] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Transactions on Information Theory* 52 (2) (2006) 489–509.
- [17] E. J. Candès, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, *IEEE Transactions on Information Theory* 52 (12) (2006) 5406–5425.
- [18] D. L. Donoho, Compressed sensing, *IEEE Transactions on Information Theory* 52 (4) (2006) 1289–1306.
- [19] R. Marks, Introduction to Shannon sampling and interpolation theory, Springer Science & Business Media, 2012.
- [20] X. Yang, H. Lei, N. A. Baker, G. Lin, Enhancing sparsity of hermite polynomial expansions by iterative rotations, *Journal of Computational Physics* 307 (2016) 94–109.
- [21] P. G. Constantine, E. Dow, Q. Wang, Active subspace methods in theory and practice: applications to kriging surfaces, *SIAM Journal on Scientific Computing* 36 (4) (2014) A1500–A1524.
- [22] J. D. Jakeman, M. S. Eldred, K. Sargsyan, Enhancing ℓ_1 -minimization estimates of polynomial chaos expansions using basis selection, *Journal of Computational Physics* 289 (2015) 18–34.
- [23] I. Babuska, R. Tempone, G. E. Zouraris, Galerkin finite element approximations of stochastic elliptic partial differential equations, *SIAM Journal on Numerical Analysis* 42 (2) (2004) 800–825.
- [24] D. Xiu, G. E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, *SIAM journal on scientific computing* 24 (2) (2002) 619–644.

- [25] M. K. Deb, I. M. Babuška, J. T. Oden, Solution of stochastic partial differential equations using galerkin finite element techniques, *Computer Methods in Applied Mechanics and Engineering* 190 (48) (2001) 6359–6372.
- [26] P. G. Constantine, M. S. Eldred, E. T. Phipps, Sparse pseudospectral approximation method, *Computer Methods in Applied Mechanics and Engineering* 229 (2012) 1–12.
- [27] S. Hosder, R. W. Walters, M. Balch, Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables, in: *Proceedings of the 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, AIAA paper, Vol. 1939, 2007.
- [28] M. Lustig, D. Donoho, J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid mr imaging, *Magnetic resonance in medicine* 58 (6) (2007) 1182–1195.
- [29] J. L. Paredes, G. R. Arce, Z. Wang, Ultra-wideband compressed sensing: channel estimation, *IEEE Journal of Selected Topics in Signal Processing* 1 (3) (2007) 383–395.
- [30] J. H. Ender, On compressive sensing applied to radar, *Signal Processing* 90 (5) (2010) 1402–1414.
- [31] C. R. Berger, Z. Wang, J. Huang, S. Zhou, Application of compressive sensing to sparse channel estimation, *Communications Magazine, IEEE* 48 (11) (2010) 164–174.
- [32] J. F. Gemmeke, B. Cranen, L. Boves, Compressive sensing for missing data imputation in noise robust speech recognition, *IEEE Journal of Selected Topics in Signal Processing* 4 (2) (2010) 272–287.
- [33] R. G. Baraniuk, Compressive sensing, *IEEE signal processing magazine* 24 (4).
- [34] E. J. Candès, Compressive sampling, in: *Proceedings of the international congress of mathematicians*, Vol. 3, Madrid, Spain, 2006, pp. 1433–1452.
- [35] E. J. Candès, M. B. Wakin, An introduction to compressive sampling, *Signal Processing Magazine, IEEE* 25 (2) (2008) 21–30.
- [36] J. Hampton, A. Doostan, Compressive sampling of polynomial chaos expansions: convergence analysis and sampling strategies, *Journal of Computational Physics* 280 (2015) 363–386.
- [37] L. Yan, L. Guo, D. Xiu, Stochastic collocation algorithms using ℓ_1 minimization, *International Journal for Uncertainty Quantification* 2 (3).
- [38] X. Yang, G. E. Karniadakis, Reweighted ℓ_1 minimization method for stochastic elliptic differential equations, *Journal of Computational Physics* 248 (2013) 87–108.
- [39] J. Peng, J. Hampton, A. Doostan, A weighted ℓ_1 minimization approach for sparse polynomial chaos expansions, *Journal of Computational Physics* 267 (2014) 92–111.
- [40] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, *SIAM review* 43 (1) (2001) 129–159.

- [41] D. L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, *Proceedings of the National Academy of Sciences* 100 (5) (2003) 2197–2202.
- [42] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, *IEEE Transactions on Information Theory* 49 (12) (2003) 3320–3325.
- [43] Z. Xu, X. Chang, F. Xu, H. Zhang, $\ell_{\frac{1}{2}}$ regularization: A thresholding representation theory and a fast solver, *IEEE Transactions on Neural Networks and Learning Systems* 23 (7) (2012) 1013–1027.
- [44] P. Yin, Y. Lou, Q. He, J. Xin, Minimization of ℓ_{1-2} for compressed sensing, *SIAM Journal on Scientific Computing* 37 (1) (2015) A536–A563.
- [45] D. L. Donoho, M. Elad, V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Transactions on Information Theory* 52 (1) (2006) 6–18.
- [46] T. T. Cai, L. Wang, G. Xu, New bounds for restricted isometry constants, *IEEE Transactions on Information Theory* 56 (9) (2010) 4388–4394.
- [47] M. A. Maleki, D. L. Donoho, R. Gray, A. Montanari, Approximate message passing algorithms for compressed sensing, Stanford University, 2010.
- [48] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *Signal Processing, IEEE Transactions on* 54 (11) (2006) 4311–4322.
- [49] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: *Proceedings of the 26th annual international conference on machine learning, ACM*, 2009, pp. 689–696.
- [50] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, F. R. Bach, Supervised dictionary learning, in: *Advances in neural information processing systems*, 2009, pp. 1033–1040.
- [51] V. Cevher, A. Krause, Greedy dictionary selection for sparse representation, *IEEE Journal of Selected Topics in Signal Processing* 5 (5) (2011) 979–988.
- [52] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, *Probabilistic Engineering Mechanics* 25 (2) (2010) 183–197.
- [53] E. Van Den Berg, M. Friedlander, Spg11: A solver for large-scale sparse reconstruction (2007).
- [54] E. Van Den Berg, M. P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, *SIAM Journal on Scientific Computing* 31 (2) (2008) 890–912.
- [55] E. J. Candes, Y. C. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, *Applied and Computational Harmonic Analysis* 31 (1) (2011) 59–73.
- [56] K. Sargsyan, C. Safta, H. N. Najm, B. J. Debusschere, D. Ricciuto, P. Thornton, Dimensionality reduction for complex models via Bayesian compressive sensing, *International Journal for Uncertainty Quantification* 4 (1).

- [57] A. Blumer, A. Ehrenfeucht, D. Haussler, M. K. Warmuth, Occam's razor, Readings in machine learning (1990) 201–204.
- [58] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Mathematics and computers in simulation 55 (1) (2001) 271–280.
- [59] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering & System Safety 93 (7) (2008) 964–979.
- [60] E. J. Candes, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, Journal of Fourier analysis and applications 14 (5-6) (2008) 877–905.